BOTTOM-UP FINDS SOME SQUARES FASTER

LEONARD GORDON
Tucson, Arizona

"Designing a List For Word Squares" in the August 1996 Word Ways shows that a word list intended for finding squares can be improved by balancing letter frequencies. In general, we want the fraction of words beginning or ending with any letter to equal the fraction at which that letter exists in the word-stock. In particular, we want to enhance the fraction of words beginning with e,i,n or o (and to a limited extent beginning with a or u), and the fraction of words ending with a,i, o or u.

The earlier article confirms Ted Clarke's and Richard Sabey's separate observations that solving straight-down is faster than straight-up when making exhaustive searches with a computer. I will now show that the old formists' bottom-up strategy works if we have the computer emulate the way they worked. After making the various tests described below, I believe that the method may well be the best for trying to find 9x9 or 10x10 squares with a computer. The argument that has been going on in Word Ways recently is due to confusing this method with straight-up exhaustive search. (Computer scientists are prone to this kind of error.)

Here is a short study on 8x8 squares. For each of two tests, I made a "gut" list. This consisted of a fraction of the Official Scrabble Players Dictionary (OSPD) balanced by eliminating some of the words beginning with b,c,m or p (and in one case, s) and then adding e,i,n words from the remaining fraction of the OSPD and from Webster's Second Unabridged (Web 2). Word endings were not balanced. Instead, I made a "bottom" list with words chosen from my entire Web 2-OSPD database. Bottom words were chosen by a method like that described by Chris Long in "Mathematics of Square Construction" in February 1993. One bottom list (arranged with the best words first) was used for both tests.

For a search, only gut words were put in the computer's lower memory. Bottom words were put on a RAM disk (a hard disk would have been OK) and read in sequentially during the search. Results were spectacular as shown in the following tables. Both tests were run for at least an hour past the time when the last reported square was found; even then, run times here are significantly shorter than the 900 minutes needed for an exhaustive search of a comparable balanced list described in the earlier article. However, the search was neither fast nor exhaustive. The method simply forced squares to appear early. The final column in the first two tables (wpm) is the average rate of testing bottom words up to that point. Note the accelerating rate; better bottom words take longer to evaluate. The third table compares times to reach squares

found in both tests. Assuming that the ratio of 1.44 is a fair measure, increasing gut list size 10 per cent increases search time 44 per cent.

Why not make a gut list and a top list and search top-down? My feeling is that it will not work very well. It has nothing to do with English being ending-poor or the number of beginning or ending bigrams. The most important factor is that the bottom list can be made to balance the disproportionate number of words ending in s,d,y. However, I am uneasy about having eliminated so many words beginning with b,c,m,p in order to improve the fraction of e,i,n beginnings. It might be wise to make a separate search using the eliminated words as a top list.

TEST #1, 23,589 GUT WORDS

	bottom word	position in list	min	wpm
1	relessee	14	38	0.37
2	strasses	21	53	0.40
3	recessed	45	85	0.53
4	sateless	153	185	0.83
5	seedages	163	193	0.85
6	sledgers	421	344	1.22
7	decedent	562	406	1.38
8	edestans	640	432	1.48
9	seedings	914	531	1.72
10	sternads	1031	563	1.83
11	decadent	2715	948	2.86
12,13	encenter	2924	971	3.01
14	aerocyst	4592	1224	3.75
15,16	ecostate	5352	1294	4.14

TIME TO FIND CERTAIN SQUARES

bottom word	minutes test#1	minutes test#2	ratio
relessee	38	67	1.76
recessed	85	136	1.60
seedages	193	289	1.50
edestans	432	623	1.44
seedings	531	767	1.44

TEST #2, 25,973 GUT WORDS

	bottom word	position in list	min	wpm
1	destress	9	51	0.18
2	relessee	14	67	0.21
3	seedness	16	77	0.21
4,5,6	seedsmen	17	78	0.22
7,8,9	sereness	19	85	0.22
10	deedless	28	97	0.29
11	recessed	45	136	0.33
12	sternest	66	161	0.41
13,14	tresslet	74	172	0.43
15,16	dressage	99	203	0.49
17	sarsenet	151	276	0.55
18	seedages	163	289	0.56
19	sledders	180	299	0.60
20	eagerest	271	383	0.71
21,22	esthesia	279	390	0.72
23	reseated	359	459	0.78
24	resected	360	460	0.78
25	sengreen	398	488	0.82
26	smartest	424	505	0.84
27,28	stayless	438	524	0.84
29-31	edestans	640	623	1.03
32,33	herseems	737	670	1.10
34	seedings	914	767	1.19